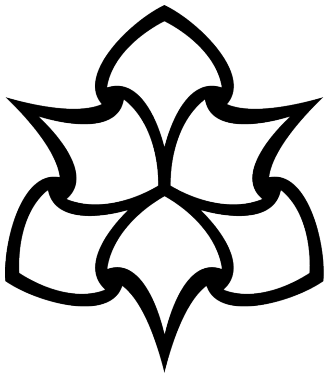

Differentially Private High-Dimensional Data Publication in Internet of Things

Zhigao Zheng, *Member, IEEE,* Tao Wang †, *Member, IEEE,*
Jinming Wen †, Shahid Mumtaz, Ali Kashif Bashir and Sajjad Hussain Chauhdary

***Abstract*—Internet of Things and the related computing paradigms, such as cloud computing and fog computing, provide solutions for various applications and services with massive and high-dimensional data, while produces threatens on the personal privacy. Differential privacy is a promising privacy-preserving definition for various applications and is enforced by injecting random noise into each query result such that the adversary with arbitrary background knowledge cannot infer sensitive input from the noisy results. Nevertheless, existing differentially private mechanisms have poor utility and high computation complexity on high-dimensional data because the necessary noise in queries is proportional to the size of the data domain, which is exponential to the dimensionality. To address these issues, we develop a compressed sensing mechanism (CSM) that enforces differential privacy on the basis of the compressed sensing framework while providing accurate results to linear queries. We derive the utility guarantee of CSM theoretically. An extensive experimental evaluation on real-world datasets over multiple fields demonstrates that our proposed mechanism consistently outperforms several state-of-the-art mechanisms under differential privacy.**

***Index Terms*—Internet of Things, compressed sensing, differential privacy, high-dimensional data, synopsis, utility.**

I. Introduction

With the advancement of Internet of Things (IoT) and data capture technologies, an unprecedented volume and variety of data are genereted constantly, and comprehensive information recording about individuals are becoming increasingly easy. An emerging wave of IoT services and applications, for exmaple, smart grids, smart healthcare, and location-based services (LBS), increasingly rely on accurate and complete data analysis, which drives much more types of data to be measured. Meanwhile, some advanced computing paradigms, for example, cloud computing [1] and fog computing [3] (a new paradigm that extends the cloud computing to the edge of the network), offer powerful data processing and analyzing plateforms for such data. Moreover, the number of attributes is increasing, thus sharply elevating the data dimension. Such high-dimensional data are of significant prevalence in many data mining and academic research domains [2], [33]. Nevertheless, reliable measures must be adopted to prevent threats to the privacy of arbitrary institutions and individuals while retaining the applicability of the published data to the abovementioned purposes because individual data contain sensitive information. A promising solution to this problem is to sanitize data before sharing them such that the inference of sensitive information by adversaries is resisted while retaining the statistical properties of the high-dimensional data [4][5].

Numerous privacy mechanisms have been proposed for privacy-preserving data publication. Among them, differential privacy is an insightful and influential privacy definition [6] that guarantees individuals' privacy when releasing the statistical information of sensitive data regardless of the arbitrary background knowledge of the adversary. Informally, differential privacy states that the deletion, addition, or modification of a single record in datasets or results by injecting random noise whose magnitude is controlled by a user-specified parameter (privacy budget) has a trivial effect on statistical query results. Hence, differential privacy can guarantee the privacy of an individual if given access to the sanitized data about all subjects but the individual. In this situation, an adversary cannot determine the individual's private value.

A plethora of variations and adaptations of differential privacy has been proposed for low-dimensional data publication with different types of input databases and application domains [8][9][18]. Nevertheless, when the input dataset has high dimensionality and many attributes, such as set-valued data [10], existing differentially private solutions must inject a prohibitive amount of noise, thereby failing to provide useful results. We focus on linear counting queries, which are essential operations in various analytical tasks. A naive differentially private mechanism injects i.i.d. noise to a query result or an attribute value with a variance of $2\Delta^2/\epsilon^2$ using the Laplace mechanism, where Δ is the sensitivity of the query. If the input dataset of the counting query poses d attributes (i.e., dimensions) and r values in each attribute, then the size of the entire output domain is r^d, which is excessively

The research work reported in this paper is supported by the National Natural Science Foundation of China (No. 61861042 and 61701453), the Fundamental Research Funds for the Central Universities (the China University of Geosciences (Wuhan), No. CUG190607, and Wuhan University), the Natural Science Foundation of China (No. 41571426), and Wuhan Applied Basic Research Program (No. 2017010201010114).

†: Corresponding author.

Zhigao Zheng is with School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China (e-mail: zhengzhigao@hust.edu.cn)

Tao Wang is with the School of Educational Information Technology, Central China Normal University, Wuhan, 430079, China. (e-mail: tmac@mail.ccnu.edu.cn)

Jinming Wen is with College of Information Science and Technology, Jinan University, Guangzhou, 510632, China. J. Wen was partially supported by NSFC (No. 11871248), Key Program of NSFC (No. 61932010) and "the Fundamental Research Funds for the Central Universities" (No. 21618329). (e-mail: jinming.wen@mail.mcgill.ca)

Shahid Mumtaz is with Instituto de Telecomunicaces, Lisboa, Portugal. (e-mail: Dr.shahid.mumtaz@ieee.org)

Ali Kashif Bashir is with Department of Computing and Mathematics, Manchester Metropolitan University, UK. (e-mail: Dr.alikashif.b@ieee.org)

Sajjad Hussain Chauhdary is with Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudia Arabia. (e-mail: shussain1@uj.edu.sa)

high, and Δ increases in a high-dimensional dataset, thereby requiring an increased magnitude of noise. Moreover, sensitivity Δ increases with the probability for batch counting queries. Executing such a query set on a high-dimensional dataset will introduce added errors and severely degrade utility. Therefore, existing solutions for low-dimensional data are either inefficient or ineffective and suffer from the curse of dimensionality.

Efforts have been exerted in attempts to solve the problem, which include decomposing high-dimensional data into a set of low-dimensional marginal tables, along with an inference mechanism that infers the joint data distribution from low-dimensional output. PriView [11], PrivBayes [12][13], sampling-based testing framework [14], DPCopula [15], DPSence [16], and DPPro [17] are representative differential privacy solutions for releasing high-dimensional data. DPPro projects a high-dimensional dataset to a randomly selected low-dimensional subspace to preserve pairwise $\mathcal{L}_2$ distances and relevant user segmentation, thereby minimizing the magnitude of noise injection while enforcing differential privacy regardless of the background knowledge of the adversary and the underlying distribution of the dataset. However, these mechanisms demand an inference mechanism and possess comparatively high computation complexity.

The idea behind another class of solutions, which includes random (linear or affine) transformation [19], Fourier transformation [7], [20], wavelet transform [21], hierarchical trees [4], compressive mechanism [22], and principal component analysis [23], is to establish a synopsis of the original high-dimensional dataset with a small size. Through releasing a noisy synopsis under differential privacy, these mechanisms can answer an arbitrary number of linear queries while maintaining good utilization by reducing the magnitude of noise necessary to satisfy the differential privacy of the dataset. These solutions offer significant information for our work. However, most of these solutions effectively respond to multi-dimensional continuous data only.

To sum up, two fundamental challenges should be addressed to achieve the secure and efficient differentially private high-dimensional data publication. First, the sensitivity function should be calculated to determine the size of the injected noise, particularly when the data dimensionality is relatively high. Second, the data utility should be ensured and the privacy-preserving mechanism cannot cause an apparent influence on query outputs. In this paper, we propose the compressed sensing mechanism (CSM) under differential privacy, which leverages the compressed sensing framework, an universal data compression scheme, to reduce the data dimensionality, so that CSM can promise ϵ-differential privacy while maintaining a high utility.

Our key contribution is a novel universal solution for publishing high-dimensional data under differential privacy. More specifically,

- we design a compressed sensing mechanism to reduce data dimensionality while enforcing differential privacy, which is built on the universal comrpressed sensing theory.
- we propose a novel sensitivity definition, sensing sensitivity, for properly determining the amount of noise for each measurement coeffcient, and we further theoretically analyze the utility of CSM under an active utility measure, namely (α, η)-usefulness.
- we conduct extensive experiments over four high-dimensional datasets for two different types of queries to evaluate the performance of CSM comprehensively. We demonstrate that CSM is superior by orders of magnitude to several state-of-the-art solutions in terms of result accuracy.

The rest of this paper is organized as follows. Section II reviews and summarizes previous studies on differentially private mechanisms for high-dimensional data publication. Section III describes notations and preliminaries. Section IV-A presents the CSM framework under ϵ-differential privacy and analyzes the privacy and utilization of CSM. The superiority of the CSM mechanism is demonstrated in Section V and through extensive experiments on real datasets, and the conclusions and suggestions for future work are in Section VI.

II. Related Works

Differential privacy was formally proposed by Dwork et al. [25], and numerous previous works have been designed in the manner of differential privacy for the publication of low-dimensional data. Differential privacy has two key advantages. First, differential privacy defines the maximum background knowledge that the adversary knows all the information about the individuals, except one sensitive record. Consequently, the adversary is powerful and can launch an arbitrary privacy attack; nevertheless, differential privacy can still guarantee the privacy of the individual. Second, differential privacy is built on a statistical probability model and thus can quantitatively analyze the risk of privacy disclosure. Therefore, differential privacy is an influential privacy definition and has substantial research value, thereby drawing significant attention in the disciplines of computer science, database systems, data mining, and machine learning. These fields generally utilize the Laplace mechanism [6], which is the basic implementation mechanism, to enforce ϵ-differential privacy, which is the original definition of differential privacy.

The extensive applications of high-dimensional data drive the growing research on differentially private high-dimensional data release. Mohammed et al. [5] presented the use of probabilistic generalization to eliminate the curse of dimensionality, which compounds rapidly at high dimensionality. Xiao et al. [26] proposed the DPCube, which is based on KD-tree partitioning, for high-dimensional healthy data; this mechanism generates a differentially private cell histogram by partitioning the noisy cell histogram mixed with Laplace noise. However, the high level of partitioning and the skewed distribution of each distribution increase the errors of perturbation and estimation, respectively, due to the large attribute domain in high-dimensional data. Qardaji et al. [11] investigated a mechanism for binary data, PriView, which uses a covering design to select a group of low-dimensional marginal tables as views and produces k-way marginal ones on the basis of

maximum entropy optimization. Zhang et al. [12] proposed PrivBayes, which iteratively learns the parent sets of the attributes in a Bayesian network by applying an exponential mechanism with a surrogate function for mutual information. The performance of PrivBayes is greatly susceptible to the randomly selected initial attribute and requires the parent sets of the attributes to have identical sizes. Similarly, Su et al. [27] developed DP-SUBN, which is based on PrivBayes; this mechanism explores a non-overlapping covering design to produce two-way marginal tables of a given set of attributes to enhance the adaptability of the Bayesian network and reduce communication cost. Li et al. [15] proposed a differentially private mechanism called DPCopula for multi- and high-dimensional data; it uses the copula function to generate a multivariate joint distribution by describing the dependencies between multivariate random vectors. Drawn on such idea, Chen et al. [14] proposed a sampling-based framework that was constructed by a generic threshold mechanism to feature a systematic inquiry on pairwise attribute dependencies and infer the joint distribution by applying the junction tree algorithm. This mechanism performs well on binary and non-binary data. However, the sampling-based inference mechanism may produce a maximum final error while minimizing the resultant error. Recently, Xu et al. [17] designed DPPro, which projects a high-dimensional dataset to a randomly chosen low-dimensional domain to preserve pairwise L-distances between individuals and address dimensionality. Thus, the magnitude of the added noise depends on the projection dimension instead of the size of the original dataset, thereby maximizing utility. Day et al. [16] presented DPSense for high-dimensional data; it uses a sensitivity control mechanism for differential privacy to publish the statistical information of the input data. Ren et al. [19] developed a local differentially private high-dimensional data publication algorithm (LoPub) by using distribution estimation techniques. Nevertheless, inferior utility and high computation complexity limit the application of DPSense. In summary, high computation complexity is a common problem among this type of differentially private mechanisms.

Several researchers have investigated the mechanisms of synopsis establishment for maximizing the utility of queries under differential privacy requirements. Rastogi et al. [20] constructed synthetic data from the original data in the Fourier domain to preserve all low-dimensional marginal data by adding Laplace noise to the discrete Fourier transformation coefficients. Then, Acs et al. [28] improved the Fourier-based mechanisms via a rigorous utility analysis. Nevertheless, the extremely large number of bins in the original histograms causes poor accuracy and a computation complexity that is proportional to the quadratic number of bins in the worst-case scenario. Privelet, which was proposed by Xiao et al. [21], is a widely adopted synopsis-based mechanism that maps multi-dimensional data to a frequency matrix and converts this matrix into a coefficient matrix via wavelet transforms. Then, Privelet adds Laplace noise to the coefficient matrix, thereby obtaining a noisy frequency matrix through inverse conversion. However, this mechanism works only for ordinal data. Hay et al. [4] proposed a hierarchical tree approach, and Cormode et al. [29] designed a statistical process for computing a private summary for sparse data without generating the entire contingency table by solely considering the scalability of the problem. The work of Li et al. [22] is the most relevant; however, it focused on the problem of privacy budget exhaustion, especially for continuous observation of datasets. However, their motivation opposed that of compressed sensing, in which reconstructing sparse data is undesirable and considered breach of privacy. These efforts benefit multi-dimensional data but pose limitations for high-dimensional data. Alternatively, these mechanisms provide significant information for designing an effective and efficient differentially private mechanism for high-dimensional data release.

III. NOTATIONS AND PRELIMINARIES

This section describes the notations and preliminaries underlying our problem. Assumptions and formal definitions are also provided.

A. Notations

Individuals intend to release a high-dimensional dataset (or table) $\mathcal{D}$ with n tuples and d distributes $\mathcal{A} = \{A_1, A_2, ..., A_d\}$, each of which is either numerical or categorical and either ordinal or nominal, respectively. The domain size of $\mathcal{D}$ is denoted as $N = \prod_{i=1}^{d} |A_i|$, which is extremely large.

The linear counting query Q is a linear combination of the statistics (counts) of the attributes in the data domain, denoted as $c_1, c_2, ..., c_d$, and it is expressed as $Q(\mathcal{D}) = q_1c_1 + q_2c_2 + ... + q_dc_d$, where q_i is the weight of the queried attribute result.

Individuals utilize a sanitization mechanism $\mathcal{M}$ to generate and publish a sanitized version of the query result $\mathcal{M}(\mathcal{D})$ to protect the privacy of $\mathcal{D}$. Table I summarizes the frequently used notations in the article.

TABLE I
SUMMARY OF FREQUENTLY USED NOTATIONS

Symbol	Description
$\mathcal{D}, \mathcal{D}'$	input datasets
$\mathcal{M}$	sanitization mechanism
ϵ, δ	privacy parameters
d, r, N	dataset parameters
Q	query function
$Q(\mathcal{D})$	exact result of Q over dataset D
$\mathcal{S}$	sanitized output results of M corresponding to Q
ξ, η	utility parameters
$\mathcal{L}$	laplace mechanism
Δ_Q, Δ_{SS}	sensitivity
Φ	Dictionary basis
$\mathcal{X}$	sparse representation of D
K	Sparsity
Ψ	measurement matrix
$\mathcal{I}$ and $\mathcal{I}^*$	measured non-noisy and noisy matrices
$\mathcal{D}^*$	noisy reconstructed dataset

B. Differential Privacy

Differential privacy is motivated by the intuition that the sanitized output generated by the input of a database is approximately indistinguishable from that generated by the input of its neighbor database. A pair of datasets, $\mathcal{D}$ and $\mathcal{D}'$, is called neighbor datasets $if f$ $\mathcal{D}'$ can be produced by adding, removing, or modifying exactly one tuple from $\mathcal{D}$.

Definition 1. *(ϵ-differential privacy). A sanitization mechanism $\mathcal{M}$ satisfies ϵ-differential privacy if it holds for any pair of neighbor datasets $\mathcal{D}$ and $\mathcal{D}'$ that*

$$Pr(\mathcal{M}(\mathcal{D}) \in S) \leq e^{\epsilon} Pr(\mathcal{M}(\mathcal{D}') \in \mathcal{S}) \tag{1}$$

where $\mathcal{S}$ denotes all possible outputs of $\mathcal{M}$ and ϵ is the privacy budget, which is mainly restricted by $\mathcal{M}$.

The inequality indicates that an adversary can possess a narrow confidence for inferring the either/or input dataset from $\mathcal{D}$ and $\mathcal{D}'$ (that is, the presence or absence of exactly one tuple in the input dataset) only by observing regardless of the adversary's background knowledge. Consequently, differential privacy guarantees the privacy of any individual with sensitive attributes in the dataset.

In practical applications, ϵ-differential privacy is generally enforced by a fundamental mechanism, the Laplace mechanism, which relies on the important parameter of L_1sensitivity.

Definition 2 (L_1-sensitivity)**.** *Given a query function Q, its L_1-sensitivity Δ_Q is the maximum L_1 distance between the results of Q over any pair of neighbor datasets $\mathcal{D}$ and $\mathcal{D}'$; it is denoted as*

$$\Delta_Q = \max_{\mathcal{D},\mathcal{D}'} \|Q(\mathcal{D}) - Q(\mathcal{D}')\|_1 \tag{2}$$

where Δ_Q is characterized by the query function Q and its output domain rather than the input dataset $\mathcal{D}$.

L_1-sensitivity underlies the Laplace mechanism, which is formally given by Definition 3.

Definition 3. *(Laplace mechanism). Given dataset $\mathcal{D}$ and query function Q, the Laplace mechanism obtains sanitized outputs $\mathcal{S}$ by injecting i.i.d. Laplace noise $\mathcal{L}$ to the exact query result with a mean of 0 and scale $\lambda = \Delta_Q/\epsilon$ and is thus defined as $\mathcal{M}_{\mathcal{L}}(\mathcal{D}) = Q(\mathcal{D}) + \mathcal{L}$.*

The variance of the added Laplace noise $\mathcal{L}$ to a query result or an attribute value is $2\Delta_Q^2/\epsilon^2$, and the overall expected squared error for Q, obtained by $\mathcal{M}_{\mathcal{L}}(\mathcal{D})$, is $2d\Delta_Q^2/\epsilon^2(2r^d\Delta_Q^2/\epsilon^2)$ when each attribute has only one value (when each attribute has r values). Excessive amounts of independent Laplace noise are bound to be added into a high-dimensional dataset.

C. Compressed Sensing

This article focuses on establishing synopsis of the original high-dimensional dataset via compressed sensing (CS). The entirely personalized and customizable data processing framework, which samples and compresses the original data by selecting the best-matched domains of sparse transform and compressed projection, accurately reconstructs the original data from the measured data of small size. This section provides a brief description of CS. The theory was discussed in detail in previous studies [30][31].

A major premise of CS is that the data are sparse or compressible, which is not always true. To overcome this obstacle, CS converts the original data into sparse data through sparse representation, an operation that seeks few vectors from a dictionary basis to represent the entire information of the original data. The sparse representation of any given dataset $\mathcal{D} \in \mathbb{R}^{d\times n}$, which is denoted as a $d \times n$ matrix, is

$$\mathcal{D} = \Phi\mathcal{X}, s.t. \|x_i\| \leq K \& i \in [1, n], \tag{3}$$

where $\Phi \in \mathbb{R}^{d\times n}$ is the dictionary basis, which is either orthogonal $d = m$ or non-orthogonal $d \neq m$; $\mathcal{X} = [x_1, x_2, ..., x_n] \in \mathbb{R}^{m\times n}$is the sparsely represented matrix of dataset $\mathcal{D}$ under the dictionary basis Φ; and $K << d$. Consequently, vector x_i is K-sparse if it has K non-zero items at most. If Φ is orthogonal, then the objective function has a unique solution, that is, $\mathcal{X} = \Phi^{-1}\mathcal{D}$. If Φ is non-orthogonal (which is a more common case), then $\mathcal{X}$ is approximately obtained by solving the following L_0-norm minimization equation:

$$\hat{\mathcal{X}} = argmin \|\mathcal{X}\|_0, s.t. \mathcal{D} = \Phi\mathcal{X} \tag{4}$$

Subsequently, by using measurement matrix $\Psi \in \mathbb{R}^{s\times d}$, a K-sparse dataset $\mathcal{D}$ can be projected into measured matrix $I \in \mathbb{R}^{s\times n}$ with a considerably reduced dimensionality, where $\mathcal{I} = \Psi\mathcal{D}$. The data projection under the measurement matrix is exactly the data compression, and the dimensionality of Ψ is substantially less than that of $\mathcal{D}$ (i.e., $s << d$). Equivalently, the mathematical equation of the entire CS process is as follows:

$$\mathcal{I} = \Psi\mathcal{D} = \Psi\Phi\mathcal{X} = \Omega\mathcal{X} \tag{5}$$

where $\Omega = \Psi\Phi \in \mathbb{R}^{s\times m}$ is the sensing matrix. Therefore, the customized data sampling and data compressing can be executed concurrently owing to the sensing matrix. The reconstruction of $\mathcal{X}$ can be rewritten as the following L_0-norm minimization:

$$\hat{\mathcal{X}} = argmin \|\mathcal{X}\|_0, s.t. I = \Omega\mathcal{X} \tag{6}$$

Candès and Tao [13] have proved that sensing matrix ΩX must fulfill the restricted isometry property (RIP) to allow the original dataset D to be reconstructed accurately.

Definition 4. *(RIP). Given a restricted isometry parameter φ_K, a sensing matrix satisfies the RIP if it holds for any sparse data that*

$$(1 - \varphi_K)\|\mathcal{X}\|_2^2 \leq \|\Omega\mathcal{X}\|_2 \leq (1 + \varphi_K)\|\mathcal{X}\|_2^2, \varphi_K \in (0, 1) \tag{7}$$

$\|\Omega\mathcal{X}\|_2^2$ and $\|\mathcal{X}\|_2^2$ are the energies of the observation and the original data, respectively, and remain unchanged after orthogonal transformation. In addition, they are the squares of L_2-norms, which are Euclidean distances from the original point. The energies of the observation and original data are approximately identical if approaches zero.

However, in real-life applications, the measured data may be corrupted by an unknown noise e and are described as

$$\mathcal{I}' = \mathcal{I} + e = \Omega\mathcal{X} + e \tag{8}$$

It has been proved that [32] the noisy sparse data $\mathcal{X}'$ reconstructed from $\mathcal{I}'$ and the expected sparse data $\mathcal{X}$ reconstructed from $\mathcal{I}$ via matching pursuit algorithms are approximately identical with a high probability.

Theorem 1. *Suppose that a sensing matrix $\Omega_{x\times n}$ satisfies RIP and the noise $\|e\| \le \theta$. Then, we hold*

$$\|\mathcal{X} - \mathcal{X}'\| \le \frac{C\,\|\mathcal{X} - \mathcal{X}_K\|_1}{\sqrt{K}} + C'\theta \tag{9}$$

for constants C and C', where $\mathcal{X}_K$ is obtained by replacing the $n-K$ coefficients with the smallest absolute value of $\mathcal{X}$ by zero.

Additionally, given sparse data $\mathcal{X}$ with magnitude $\mathcal{R}$, we indicate that $\|\mathcal{X} - \mathcal{X}_K\|_1 \le C_p \mathcal{R} K^{1-\frac{1}{p}}$ for the constant $p \in (0,1)$. Then, the measured data $\mathcal{I}'$ can be reconstructed by $\mathcal{I}' = \Omega\mathcal{X}'$, and we hold $\|\mathcal{I} - \mathcal{I}'\|_2 = \|\mathcal{X} - \mathcal{X}'\|_2$.

CS converts a dataset $\mathcal{D}$ of size $d \times n$ into a measured matrix $\mathcal{I}$ of size $s \times n$ using sensing matrix Ω. Subsequently, we add noise to measured matrix $\mathcal{I}$ instead of the original data. The added noise can diffuse in the entire dataset after the reconstruction, thereby affecting privacy preservation (e.g., satisfying differential privacy) in a nearly similar manner as do primitive methods but with a substantially smaller amount of noise.

We focus on coping with the publication of high-dimensional data due to the increasing computation complexity and deteriorating utility caused by dimensionality. We aim to answer one or a batch of linear counting queries with the most rudimentary operation in various data statistics and analytical applications and a maximum overall utility while ensuring differential privacy. Particularly, our proposed CSM uses the fundamental Laplace mechanism to enforce ϵ-differential privacy.

IV. CSM FRAMEWORK

This section presents a detailed description of our proposed CSM framework. We first illustrate the overview of CSM in Section IV-A and then provide analyses of the privacy and utility of CSM in Sections IV-B and IV-C, respectively.

A. Overview of CSM

The main idea of the proposed CSM is to exploit the dimensionality reduction property of CS to reduce the amount of noise required to satisfy differential privacy. CSM provides the expected privacy guarantee with a small privacy budget (and thus reduced noise injection), thereby providing accurate statistical query results after reconstruction. The proposed CSM regards a dataset $\mathcal{D}$ of size $d\times n$ and privacy parameters ϵ as the input to and outputs of a noisy version $\mathcal{D}^*$ of $\mathcal{D}$. Roughly, CSM has four steps, as described in figure 1.

First, CSM implements a sparse representation of $\mathcal{D}$ (that is preprocessed generally) and maps $\mathcal{D}$ to another matrix $\mathcal{X}$ via a dictionary basis Ω. If the dictionary basis is orthogonal, then each entry in $\mathcal{X}$ can be seen as a linear combination of the entries in $\mathcal{D}$, and $\mathcal{D}$ can be losslessly obtained from $\mathcal{X}$ by a linear inverse operation. Conversely, if the dictionary basis is non-orthogonal, then each entry in X can be seen as a random affine combination of the entries in $\mathcal{D}$, and $\mathcal{D}$ can be approximately obtained from $\mathcal{X}$ by greedy algorithms [13]. The entries in $\mathcal{X}$ are the sparse coefficients.

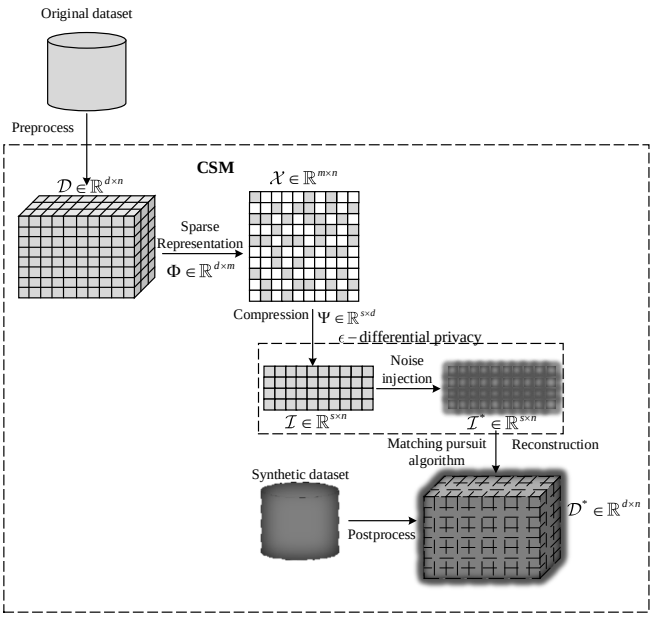

Fig. 1. CSM framework

Algorithm 1 CSM

Input: $\mathcal{D}$, ϵ, Φ, Ψ, and OMP algorithm
Output: $\mathcal{D}^*$

1: Compute $\mathcal{D} = \Phi\mathcal{X}$ to generate a sparse representation $\mathcal{X}$ via the dictionary basis Φ;
2: Project $\mathcal{X}$ into a measured matrix $\mathcal{I}$ via a measurement matrix Ψ;
3: Select noise parameter λ;
4: Acquire a noisy version $\mathcal{I}^*$ by adding the noise based on privacy parameters ϵ;
5: Reconstruct $\mathcal{X}^*$ via OMP algorithm;
6: Compute $\mathcal{D}^* = \Phi\mathcal{X}^*$
7: **return** $\mathcal{D}^*$

Second, CSM projects $\mathcal{X}$ into a measured matrix $\mathcal{I}$ via a measurement matrix Ψ. The entries in $\mathcal{I}$ are defined in this article as the measure coefficients.

Third, CSM adds i.i.d. Laplace noise to each measurement coefficient in a manner that enforces ϵ-differential privacy, thereby generating a new matrix $\mathcal{I}^* = \mathcal{I} + \mathcal{L}$ with noisy coefficients.

Finally, CSM post processes $\mathcal{I}^*$ and then converts $\mathcal{I}^*$ back to noisy dataset $\mathcal{D}^*$, which is returned as the output. The reconstruction step of CSM relies only on $\mathcal{I}^*$, thereby ensuring that CSM does not disclose any information about $\mathcal{D}$, except that in $\mathcal{I}^*$. The detailed procedure of CSM is presented in Algorithm 1.

Proposition 1. *CSM enforces ϵ-differential privacy.*

Proof. The compression process of CS can be characterized as a sanitization operation $\mathcal{K} : \mathcal{X} \to \mathbb{R}^n$. For any dataset

$\mathcal{D} \in \mathcal{X}$, $\mathcal{K}(\mathcal{D}) = \Psi\mathcal{D}$. The measurement matrix Ψ, which is generally a random matrix, is produced by sampling i.i.d. entries from a probability distribution (such as Gaussian and Bernoulli), or a matrix (such as Fourier and Hadamard). Then, the sensitivity of $\mathcal{K}$ is regarded as the distribution parameter or the number of vectors selected from the matrices. The Laplace mechanism $\mathcal{M}_{\mathcal{L}}$ embedded in the CS framework enforces ϵ-differential privacy in accordance with Definition 3. Moreover, the subsequent data reconstruction is deterministic without involving probability calculation. Therefore, we can infer that CSM enforces ϵ-differential privacy. □

B. Privacy Analysis

Intuitively, the privacy guarantee of CSM depends on the third step, in which CSM injects specific noise into the measurement coefficients of matrix $\mathcal{I}$. $\mathcal{I}$ is compressed with the dimensionality of (is considerably smaller than d.). Thus, arbitrarily changing one attribute value of one tuple or an entire tuple will change the entire measurement coefficients in the corresponding row of $\mathcal{I}$. Such changes can be concealed with the addition of an appropriate amount of noise to $\mathcal{I}$.

The response of each measurement coefficient to changes in varies, and so does the noise required for each coefficient. Drawing on the idea of Xiao et al. [21], CSM generates the amount of noise for each measurement coefficient by a magnitude function $\mathcal{H}$, which alters each coefficient to a positive real number. Accordingly, the magnitude of noise on a coefficient of $\mathcal{I}$ is $\Delta_Q/\epsilon\mathcal{H}(\mathcal{J})$, where $\mathcal{J}$ is an arbitrary measurement coefficient. Then, we provide a new sensitivity definition, that is, sensing sensitivity.

Definition 5. *(Sensing Sensitivity). Given a query function Q that inputs a matrix and outputs a real number, the sensing sensitivity Δ_{ss} of Q with respect to magnitude function $\mathcal{H}$ is expressed as*

$$\Delta_{ss} = \max_{\mathcal{D},\mathcal{D}'}(\frac{1}{\tau}\mathcal{H}(\mathcal{I})\left|Q(\mathcal{D}) - Q(\mathcal{D}')\right|) \tag{10}$$

where τ is the compression coefficient and $\tau \in (0, 1]$.

The sensing sensitivity acquires the intention of L_1-sensitivity as a special case. Specifically, for any query, the L_1-sensitivity of Q equals the sensing sensitivity with respect to $\mathcal{H}$, which assigns each measurement coefficient the same magnitude, and $\tau = 1$.

Theorem 2. *Given a query function Q with the sensing sensitivity Δ_{ss}, a sanitization mechanism $\mathcal{M}$ satisfies $2\tau\epsilon/\Delta_{ss}$-differential privacy if it holds for any $\mathcal{D}$ and $\mathcal{D}^*$ that*

$$\mathcal{M}_{CSM} = \sup_{\mathcal{D},\mathcal{D}'} ln\frac{Pr(Q(\mathcal{D}) \in \mathcal{Y})}{Pr(Q(\mathcal{D}^*) \in \mathcal{Y})} \leq \frac{2\tau\epsilon}{\Delta_{ss}} \tag{11}$$

where $\mathcal{D}^* = \mathcal{M}_{CSM}(\mathcal{D}, \mathcal{L}(\frac{\Delta_{ss}}{\epsilon}))$, and $\mathcal{Y}$ is the possible output space

Proof. Suppose $\mathcal{D}$, $\mathcal{D}'$ be any two datasets that differ in only one column vector, and correspondingly $\mathcal{I}$, $\mathcal{I}'$ be the measured matrix that also differ in only one column vector. Since Q has a sensing sensitivity Δ_{ss}, we have

$$\sum_{q\in Q}\frac{1}{\tau}\mathcal{H}(\mathcal{I})\left|Q(\mathcal{I}) - Q(\mathcal{I}*)\right|) \leq \tau\left\|\mathcal{I} - \mathcal{I}'\right\|_1 = \tau \tag{12}$$

Then, let q_j ($j \in [1, |Q|]$) be the j-th query in Q and $\mathcal{Y}$ is the possible output space. We have

$$\begin{aligned}
\mathcal{M}_{CSM} :&= \sup_{\mathcal{D},\mathcal{D}'} ln\frac{Pr(Q(\mathcal{D}) \in \mathcal{Y})}{Pr(Q(\mathcal{D}^*) \in \mathcal{Y})} \\
&= \frac{Pr\left(\mathcal{M}_{CSM}(\mathcal{D}) = [y_j]\right)}{Pr\left(\mathcal{M}_{CSM}(\mathcal{D}^*) = [y_j]\right)} \\
&= \frac{\prod_{i=1}^{|Q|}\left(\frac{\mathcal{H}(\mathcal{I})}{2\lambda}\cdot\exp\left(-\mathcal{H}(\mathcal{I})\cdot\left|y_j - q_i(\mathcal{I}^*)\right|/\lambda\right)\right)}{\prod_{i=1}^{|Q|}\left(\frac{\mathcal{H}(\mathcal{I})}{2\lambda}\cdot\exp\left(-\mathcal{H}(\mathcal{I})\cdot\left|y_j - q_i(\mathcal{I})\right|/\lambda\right)\right)} \\
&\leq \prod_{i=1}^{|Q|}\left(\exp\left(-\mathcal{H}(\mathcal{I})\cdot\left|q_i(\mathcal{I}) - q_i(\mathcal{I}^*)\right|/\lambda\right)\right) \\
&\leq 2\prod_{i=1}^{|Q|}\left(\exp\left(-\left|q_i(\mathcal{I}) - q_i(\mathcal{I}^*)\right|/\lambda\right)\right) \\
&\leq \frac{2\tau\epsilon}{\Delta_{ss}},
\end{aligned}$$

thereby proving that CSM fulfills differential privacy. □

C. Utility Analysis

In real-life applications, individuals generally cannot set the desired privacy parameters involved in differential privacy and prefer to set the intuitive utility level of the query results. A small indicates strong privacy preservation. Hence, this section analyzes the utilization of released data by the definition of (μ,η)-usefulness [24] for CSM that satisfies the given utility requirement.

Definition 6. *((μ,η)-usefulness) A mechanism $\mathcal{M}$ has (μ,η)-usefulness with respect to query function Q and dataset $\mathcal{D}$ under the $\|\bullet\|_1$-norm if it holds for parameters $\mu > 0$ and that $0 < \eta < 1$ that*

$$Pr(\|\mathcal{M}(Q,\mathcal{D}) - Q(\mathcal{D})\|_1 \geq \eta) \leq \mu \tag{13}$$

Proposition 2. *Given query Q, dataset $\mathcal{D}$, and user-specified parameters $\mu > 0$ and $0 < \eta < 1$, CSM returns $(\mu, \frac{1}{2}\exp(-\frac{\mu\epsilon}{s\Delta_{ss}}))$ -useful results of Q on $\mathcal{D}$.*

Proof. We first use $\mathcal{U}$ to represent the error introduced by CSM such that

$$Pr(\|\mathcal{M}_{CSM}(Q,\mathcal{D}) - Q(\mathcal{D})\|_1) \leq Pr(\mathcal{U} > \mu) \tag{14}$$

On the basis of the characteristics of the Laplace distribution involved in Proposition 2, we have

$$Pr(\mathcal{U} > \mu) = \int_{-\infty}^{\mu} f(x)dx = 1 - \frac{1}{2}e^{-\frac{\mu\epsilon}{s\Delta_{ss}}} \tag{15}$$

Then, we obtain

$$Pr(\mathcal{U} \leq \mu) = 1 - Pr(\mathcal{U} > \mu) = \frac{1}{2}e^{-\frac{\mu\epsilon}{s\Delta_{ss}}} \tag{16}$$

Therefore, given μ, η is expressed as

$$\eta = \frac{1}{2}\exp(-\frac{\mu\epsilon}{s\Delta_{ss}}) \tag{17}$$

□

The sensing sensitivity will be remarkably high when the queried dataset comprises several attributes. The utility of CSM is positively relevant to the dimensionality of compressed data, thereby guaranteeing a significantly lower noise level than do standard differentially private mechanisms on high-dimensional datasets.

V. EXPERIMENTS

This section experimentally evaluates the performance of CSM under -differential privacy. The privacy strength, result accuracy, and computation complexity of CSM and the impact of several core parameters are evaluated in comparison with those of state-of-the-art representative mechanisms. Laplace mechanism (LM) [6], Privlet [21], hierarchical mechanism (HM) [4], Fourier mechanism (FM) [28], PrivBays [12][13], and DPPro [17] are selected for comparison.

A. Experimental Configurations and Datasets

We perform all experiments on a desktop PC with an Intel quad-core i7-4790 @ 3.6 GHz CPU and 8 GB RAM. In each experiment, every algorithm is executed 20 times, and the average indicators are reported.

We use four datasets in our experiments [12][17], namely, AOL, Retail, UCI Adult, and TPC-E, to demonstrate CSM. These real-world datasets contribute to our evaluations and illustration of the effectiveness of our proposed mechanism in real-life applications.

AOL: This dataset is a search log that includes search keyword statistics and contains 45 different attributes after our preprocessing.

Retail: This dataset contains information about a retail market basket in which each record consists of diverse items purchased in a shopping operation. It contains 50 different attributes after our preprocessing.

UCI Adult: This dataset originally involves information about 45,222 individuals; these data were extracted from the 1994 U.S. Census and have 14 attributes of which six are continuous and eight are categorical. We consider 30,162 records after preprocessing.

TPC-E: This dataset contains the information in the "Trade", "Security", "Security status" and "Trade type" tables in the TPC-E benchmark. We summarize the statistics of the datasets in Table II.

TABLE II
DATASET CHARACTERISTICS

Datasets	Cardinality	Dimensionality	Domain Size
AOL	619,418	45	2^{45}
Retail	88,162	50	2^{50}
UCI Adult	45,222	15	2^{52}
TPC-E	40,000	24	2^{77}

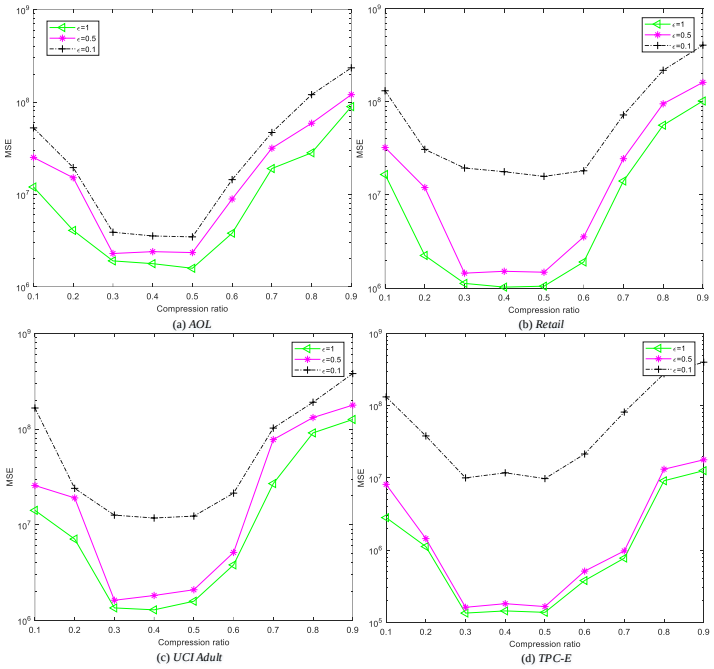

Fig. 2. Effect of s on different datasets

B. Experimental Evaluation Methodology

Over each dataset, we generate and execute 10,000 random queries, which comprise only linear counting and linear range-counting queries. For linear range-counting queries, each query Q_j sums the counts in a range $[a_i, b_i]$. Starting and end points a_i and b_i, respectively, are randomly generated and follow a uniform distribution. The utility is measured by two performance metrics, namely, (μ, η)-usefulness and mean squared error (MSE). The scalability of the mechanisms is weighed by the running time. Specifically, the MSE is the mean squared L_2 distance between the exact and noisy query answers. The experimental results offer crucial insights into the adaptation of privacy parameters for maximizing the utility of CSM.

C. Impact of Compressed Dimensionality on CSM

The compressed dimensionality s is an important parameter in CSM. It determines the dimensionality of matrices $\mathcal{I}$ and $\mathcal{I}^*$, thereby determining the necessary magnitude of noise. An excessively small s leads to increased pressure on data reconstruction, thereby degrading the computing efficiency and reconstruction accuracy, whereas an extremely large s leads to a large amount of necessary noise and consequently poor accuracy of queries. Then, we set ϵ to 1, 0.5, and 0.1, measure CSM with varying s by controlling the compression ratio (ratio of compressed dimensionality to original dimensionality) on the four datasets, and record the MSE of CSM. Fig. 2 indicates that with the compression ratio increase, the MSE of CSM first declines, and then stabilizes, and finally increases. This result is consistent with our analysis in Subsection IV.C.

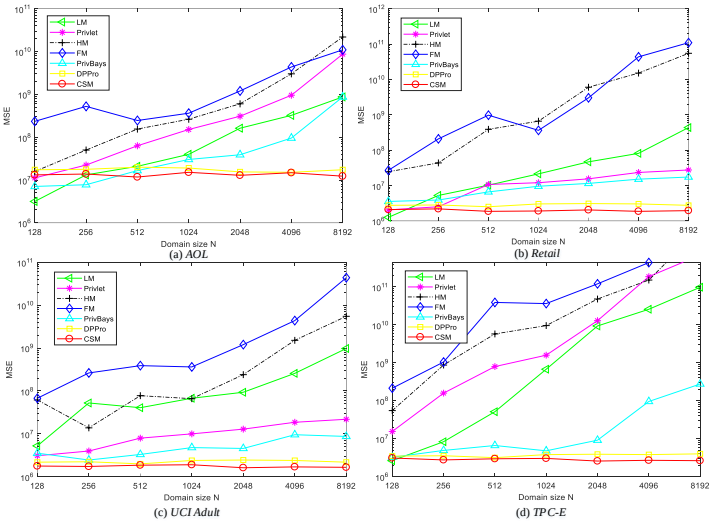

Fig. 3. Effect of domain size N on different datasets

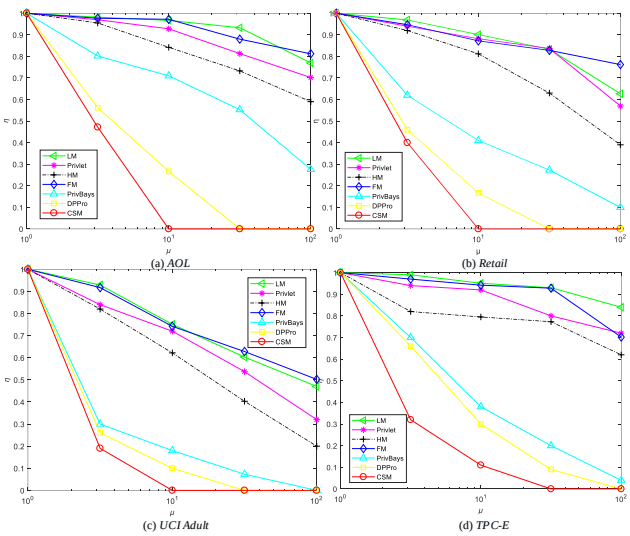

Fig. 4. (μ, η)-usefulness in different datasets

D. Impact of Varying Domain Size on CSM

We evaluate the utility of all mechanisms with the varying domain size N from 128 to 8192 at interval of 128 and fix privacy parameter ϵ to 0.1 and s to 0.3. The MSE of each mechanism in the four datasets is reported in Fig. 3. Intuitively, LM outperforms all others when the domain is relatively small partly because these counting queries are generally random and independent. Meanwhile, all other data-independent mechanisms incur an error linear to domain size N. The errors of DPPro and CSM remain the same because their errors rely on data dimensionality (after projection or compression) that is smaller than the domain size.

E. Utility Evaluation

We measure utility performance by (μ,η)-usefulness and MSE on the four datasets. In this experiment, the compressed dimensionality s and the domain size N are set to 0.4 and 256, respectively. As shown in Fig. 4, where $\epsilon = 1$, our proposed CSM has a considerably lower η than do the other mechanisms, indicating a higher probability $1-\eta$ under the same μ and ϵ, than the other mechanisms. The usefulness of CSM increases with the privacy guarantee level ϵ. Equivalently, under the same utility requirement, CSM provides a significantly preferable privacy guarantee over the other mechanisms. Hence, CSM exhibits apparent superiority of utilization and privacy over state-of-the-art mechanisms.

Fig. 5 illustrates that CSM poses a considerably lower MSE than the other mechanisms and outperforms them in answering many queries. CSM achieves the expected privacy-preserving level for high-dimensional data by introducing a small amount of noise. With the privacy parameter ϵ varying from 0.1 to 0.9, the corresponding MSE of each mechanism, except LM, decreases slowly because LM was designed without any optimization on high-dimensional data. The MSEs of the mechanisms decline as ϵ increases from 0.1 to 0.5 more notably than that with ϵ increasing from 0.5 to 0.9. This finding indicates that a large amount of noise is required to obtain a strong privacy guarantee. CSM presents a comparatively stable utility when ϵ exceeds 0.5 and accordingly indicates that CSM maintains a high-level data utility while satisfying the expected privacy requirement of individuals.

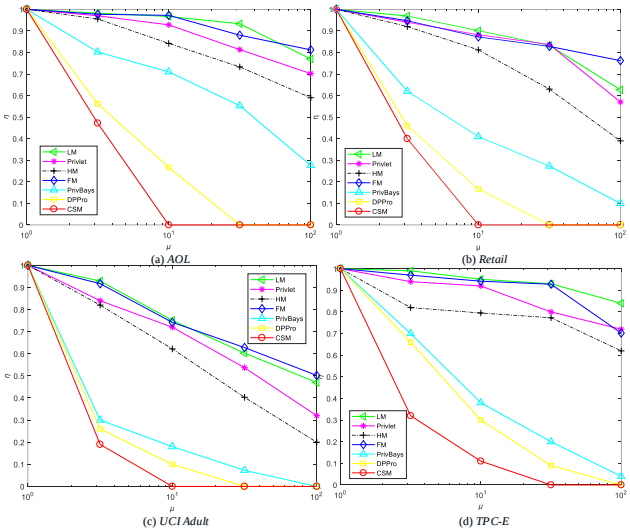

Fig. 5. MSEs in different datasets

F. Scalability of CSM

We finally demonstrate the scalability and efficiency of CSM. Fig. 6 illustrates the average running time (ART) of CSM for the two types of queries with the domain size N varying from 128 to 8192 and the number of queries Q varying from 64 to 256. Roughly, the logarithmic scale of the ART of CSM increases linearly with the logarithmic scale of the domain size. In all experimental scenarios, CSM always terminates within 17 min for each experiment, which is sufficient for achieving adequate query result accuracy.

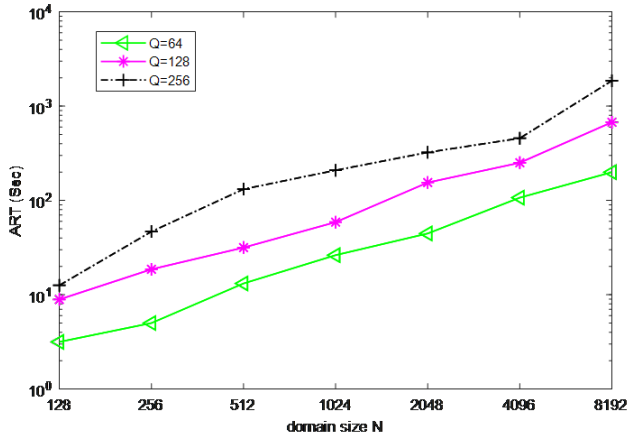

Fig. 6. MSEs in different datasets

VI. Conclusions

This article presents the compressed sensing mechanism (CSM), an optimization framework that addresses the challenges in differentially private high-dimensional data publication. It can minimize the overall error of query results under ϵ-differential privacy by injecting the minimum amount of noise into the compressed data with CS. Extensive experiments demonstrate that CSM significantly outperforms other state-of-art differentially private mechanisms for high-dimensional data publication by orders of magnitude.

The next step is to extend this work to correlated datasets with high dimensionality and large domain sizes. Given that data correlations distinctly result in high-complexity privacy-preserving mechanisms, particularly on high-dimensional datasets with large domain sizes, an inappropriate data process may generate unacceptable computation overhead. In addition, we will extend CSM to (ϵ, δ)-differential privacy.

References

[1] H. Cai, B. Xu, L. Jiang and Athanasios V. Vasilakos, “IoT-Based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges”. IEEE Internet of Things Journal, vol.4(1), pp.75-87, 2017.

[2] S.Liu, D. Maljovec, B. Wang, P.T. Bremer and V. Pascucci, “Visualizing High-Dimensional Data: Advances in the Past Decade”. IEEE Transactions on Visualization and Computer Graphics , vol.23(3), pp.1249-1268, 2017.

[3] T. Wang, J. Zhou, A. Liu, M.Z.A. Bhuiyan, G. Wang, W. Jia, “Fog-Based Computing and Storage Offloading for Data Synchronization in IoT,” IEEE Internet of Things Journal, vol.6(3), pp.4272-4282, 2019.

[4] R. McKenna, G. Miklau, M. Hay and A. Machanavajjhala, “Optimizing error of high-dimensional statistical queries under differential privacy,” Proceedings of the VLDB Endowment, vol.11, no.10, pp.1206-1219, 2018.

[5] V. Bindschaedler, R. Shokri and C. A. Gunter, "Plausible deniability for privacy-preserving data synthesis," Proceedings of the VLDB Endowment, vol.10, no.5, pp. 481-492, 2017.

[6] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy", Foundations and Trends® in Theoretical Computer Science: vol. 9, no. 3–4, pp 211-407, 2014

[7] A. Jolfaei, K. Kant, and H. Shafei, "Secure Data Streaming to Untrusted Road Side Units in Intelligent Transportation System," in Proceedings of 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering, IEEE, 2019, pp.793-798.

[8] S. Yaseen, S. M. A. Abbas, A. Anjum, T. Saba, A. Khan, S.U. R. Malik, N. Ahmed, B. Shahzad, and A. K. Bashir, "Improved Generalization for Secure Data Publishing," IEEE Access. vol. 6, pp. 27156-27165, 2018.

[9] Z. Zhang, Z. Qin, L. Zhu, J. Weng and K. Ren, “Cost-friendly differential privacy for smart meters: Exploiting the dual roles of the noise,” IEEE Transactions on Smart Grid, vol.8, no.2, pp.619-626, 2017.

[10] S. Wang, L. Huang, Y. Nie, P. Wang, H. Xu and W. Yang, “PrivSet: Set-Valued Data Analyses with Locale Differential Privacy,” in Proceedings of IEEE Conference on Computer Communications (INFOCOM’18), IEEE, 2018, pp. 1088-1096.

[11] W. Qardaji, W. Yang and N. Li, “Priview: practical differentially private release of marginal contingency tables,” in Proceedings of 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 1435-1446

[12] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava and X. Xiao, “Privbayes: Private data release via bayesian networks,” in Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014, pp. 1423-1434.

[13] S. Yao, A. K. Sangaiah, Z. Zheng and T. Wang, “Sparsity estimation matching pursuit algorithm based on restricted isometry property for signal reconstruction”, Future Generation Computer Systems, vol.88, pp.747-754, 2018..

[14] R. Chen, Q. Xiao, Y. Zhang and J. Xu, “Differentially private high-dimensional data publication via sampling-based inference,” in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 129-138.

[15] H. Li, L. Xiong and X. Jiang, “Differentially private synthesization of multi-dimensional data using copula functions,” in Advances in database technology: proceedings. International Conference on Extending Database Technology, NIH Public Access, 2014, pp. 475.

[16] W. Y. Day and N. Li, “Differentially private publishing of high-dimensional data using sensitivity control,” in Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, ACM, 2015, pp. 451-462

[17] C. Xu, J. Ren, Y. Zhang, et al., “DPPro: Differentially Private High-Dimensional Data Release via Random Projection,” IEEE Transactions on Information Forensics and Security, vol.12, no.12, pp.3081-3093, 2017.

[18] S. Ghane, A. Jolfaei, L. Kulik, and K. Ramamohanarao, "Differentially Private Streaming to untrusted Edge Servers in Intelligent Transportation System," in Proceedings of 18th IEEE International Conference On Trust, Security and Privacy In Computing and Communications/13th IEEE International Conference On Big Data Science and Engineering, IEEE, 2019, pp. 781-786.

[19] X. Ren, C. M. Yu, W. Yu, S. Yang, X. Yang, J. A. McCann and P. S. Yu, "LoPub: High-Dimensional Crowdsourced Data Publication With Local Differential Privacy," IEEE Transactions on Information Forensics and Security, vol.13, no.9, 2151-2166 ,2018.

[20] V. Rastogi and S. Nath, “Differentially private aggregation of distributed time-series with transformation and encryption,” in Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD’10), 2010, pp. 735–746.

[21] X. Xiao, G. Wang and J. Gehrke, “Differential privacy via wavelet transforms,” IEEE Transactions on Knowledge and Data Engineering, vol.23, no.8, pp.1200-1214, 2010.

[22] Y. D. Li, Z. Zhang, M. Winslett and Y. Yang, “Compressive mechanism: Utilizing sparse representation in differential privacy”, in Proceedings of The 10th Annual ACM Workshop on Privacy in The Electronic Society, ACM, 2011, pp. 177-182

[23] W. Jiang, C. Xie and Z. Zhang, “Wishart Mechanism for Differentially Private Principal Components Analysis,” in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016, pp. 1730-1736.

[24] A. Blum, K. Ligett and A. Roth, “A learning theory approach to non-interactive database privacy,” Journal of the ACM (JACM), vol. 60, no.2, pp.12, 2013.

[25] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Foundations and Trends in Theoretical Computer Science, vol.9, no.3-4, pp.211-407, 2014.

[26] Y. Xiao, J. Gardner and L. Xiong, “Dpcube: Releasing differentially private data cubes for health information,” in Proceedings of 2012 IEEE 28th International Conference on Data Engineering (ICDE’12), 2012, pp. 1305-1308.

[27] S. Su, P. Tang, X. Cheng, R. Chen and Z. Wu, ”Differentially private multi-party high-dimensional data publishing,” in Proceedings of IEEE 32nd International Conference on Data Engineering (ICDE) ,2016, pp. 205-216.

[28] G. Acs, C. Castelluccia and R. Chen, “Differentially private histogram publishing through lossy compression,” in Proceedings of IEEE 12th International Conference on Data Mining (ICDM), 2012, pp. 1-10

[29] G. Cormode, C. Procopiuc, D. Srivastava, and T. T. Tran, “Differentially private summaries for sparse data,” in Proceedings of the 15th International Conference on Database Theory, ACM, 2012, pp. 299-311.

[30] C.A. Metzler, A. Maleki and R. G. Baraniuk, ”From Denoising to Compressed Sensing,” IEEE Transactions on Information Theory, vol.62, no.9, pp.5117-5144, 2016.

[31] M. Azhar, H. Dawood, H. Dawood, G. I. Choudhry, A. K. Bashir, and S. H. Chaudhary, "Detail-preserving switching algorithm for the removal of random-valued impulse noise," Journal of Ambient Intelligence and Humanized Computing, early access, pp.1-21, 2018.

[32] M. Mangia, A. Marchioni, F. Pareschi, R. Rovatti, R. Rovatti and R. Setti, "Chained Compressed Sensing: A Block-Chain-inspired Approach for Low-cost Security in IoT Sensing," IEEE Internet of Things Journal, Early Access, 1-1, 2019.

[33] A. Jolfaei, and K. Kant, "Privacy and Security of Connected Vehicles in Intelligent Transportation System," in Proceedings of 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks–Supplemental Volume (DSN-S), IEEE, 2019: 9-10.